

[19]中华人民共和国国家知识产权局

[51]Int. Cl⁶

H04L 12/56

H04Q 11/04

[12] 发明专利申请公开说明书

[21] 申请号 97180142.8

[43]公开日 1999 年 12 月 15 日

[11]公开号 CN 1238875A

[22]申请日 97.10.27 [21]申请号 97180142.8

[30]优先权

[32]96.10.29 [33]FI [31]964354

[86]国际申请 PCT/FI97/00655 97.10.27

[87]国际公布 WO98/19428 英 98.5.7

[85]进入国家阶段日期 99.5.27

[71]申请人 诺基亚电信公司

地址 芬兰埃斯波

[72]发明人 玛第·赫尔科宁

[74]专利代理机构 中国国际贸易促进委员会专利商标事
务所

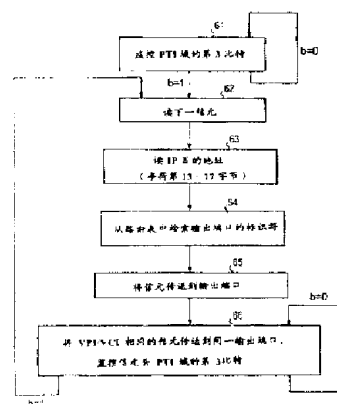
代理人 张 维

权利要求书 2 页 说明书 9 页 附图页数 5 页

[54]发明名称 电信系统中包的路由寻址

[57]摘要

本发明涉及利用路由器传送包的包交换电信网络中的包传送。需要传送的数据包被分段成较低协议层数据单元,在网络中的路由器间传送。为了以尽可能简单的方式加快包传送,(a)根据到来的数据单元(50),监控两个连续数据包(30)之间边界的出现,(b)在检测到数据包开始之后,从数据单元中读取该数据包的目的地址,(c)搜索单个数据单元的与目的地址相关联的输出端口,将该数据单元寻路到该输出端口,(d)只要连接标识符与所述数据单元相同的所有其它数据单元的内容表明数据包仍在继续,则直接根据所述多个数据单元所包含的连接标识符进行这些数据单元的路由寻址。



ISSN 1008-4274

权 利 要 求 书

1. 一种在利用路由器 (RT1...RT6) 传送数据包的包交换电信网络中传送数据包的方法, 按照该方法, 将需要传送的数据包 (30) 分段成较低协议层数据单元 (50), 在网络中的路由器间传送,

其特征在于, 在网络的路由器中

- 根据到来的数据单元 (50), 监控两个连续数据包 (30) 之间边界的出现,

- 在检测到数据包开始之后, 从数据单元中读取该数据包的目的地址,

- 搜索单个数据单元的与目的地址相关联的输出端口, 将该数据单元寻路到输出端口,

- 只要连接标识符与所述数据单元相同的所有其他数据单元的内容表明数据包仍在继续, 则根据所述多个数据单元所包含的连接标识符直接进行这些所有其他数据单元的路由寻址。

2. 根据权利要求 1 的方法, 其中所述分段数据单元是 ATM 信元, 其特征在于, 基于到来信元的消息头中 PTI 域监控两个连续包间边界的出现以及包是否仍在继续。

3. 根据权利要求 2 的方法, 其中高层包是 IP 包, 其特征在于, 从包中的第一信元读取目的地址, 且仅搜索包中第一信元的输出端口。

4. 根据权利要求 3 的方法, 其特征在于, 首先从内容可编址存储器 (73) 中检索与目的地址相关联的输出端口, 其中目的地址作为输入提供给内容可编址存储器 (73)。

5. 一种在包交换电信网络中传送包的路由器装置, 在该网络中需要传送的数据包 (30) 被分段成较低协议层数据单元 (50), 在网络中的路由器间传送, 该路由器装置包括

- 第一读取装置 (70、72), 用于读取数据包的目的地址;

- 第二读取装置 (70), 用于读取到来数据单元的连接标识符;



- 监控装置（70），基于到来数据单元监控两个连续包间边界的出现；

- 确定装置（74，75），用于确定与读出的目的地址相关联的输出端口；

- 交换装置（76），用于将数据单元路由寻址到所需输出端口；其特征在在于，

- 所述第一读取装置响应于检测到每个包的开始，直接从较低协议层数据单元中读取，并且路由寻址装置还包括：

- 配置装置（71），用于在每个数据包开始时配置交换装置，使得只要该数据包仍在继续，其连接标识符与包中第一数据单元相同的数据单元就被直接寻路到与所述第一数据单元相同的输出端口。

6. 根据权利要求5的路由器装置，其中确定装置包括路由表和软件检索处理（74），其特征在在于，确定装置还包括内容可编址存储器（73），充当路由表的高速缓存。

说明书

电信系统中包的路由寻址

本发明一般涉及包交换电信系统中实现的包传送，尤其涉及通过 ATM 网络传送 IP 包（IP，因特网协议）。

IP 是当前网络层（OSI 模型中的第三层）最普遍的协议，是因特网得以飞速发展的主要原因，将来 IP 可能会越来越普及。随着连接到因特网的主机数量的指数性增长，IP 网络的吞吐量成为瓶颈，需要新方式以较目前更快地传送 IP 业务量。

图 1 说明了 IP 网络的典型结构。在办公环境中，个人计算机 PC 或类似的终端设备连接到局域网 LAN1...LAN3，这些局域网一般是以以太网。局域网则通过路由器（RT1...RT6）与骨干网（WAN，广域网）互连。同一局域网中的所有计算机具有相同的 IP 网络地址。如果从连接到局域网的一台计算机发送一个数据包，则发送计算机的协议栈的 IP 层检查 IP 目的地址是否与其自身的 IP 网络地址相同。如果地址相同，则不需要路由器，而是通过局域网将该包发送给对应该目的地址的计算机。如果目的地的 IP 网络地址与发送计算机的 IP 网络地址不同，则计算机将该包转发给路由器，后者将包进一步传送到另一网络。

路由器之间的传输链路可以由 PDH 或 SDH 技术，或者通过包网络技术，例如（ATM、帧中继、X.25）实现。

路由器主要有两个功能：包传送和更新路由表。包传送处理的工作原理是，路由器首先从到来的 IP 包中读取目的网络地址。之后它根据它的路由表找到与该地址相关联的输出端口，通过所述端口将包发送到下一路由器。包在路由器之间传送，直至一个路由器发现目的地址与其自身的网络地址相同，在这种情况下，它将该包发给目的主机。

随着传送速率需求的增长，新技术得以引入。ATM 技术作为骨

干网技术的趋势正在增长，因为它能够进行大容量骨干连接。在这种情况下，路由节点中已面向 ATM 网络构造了接口；首先根据从 ATM 网络到达的信元重构包，为包进行路由寻址，之后再次将包分解成 ATM 信元，在 ATM 网络中传送。标准的 ATM 适配层（AAL）完成 IP 包的分解和重构。下面将详细予以描述，为后面本发明的描述提供背景。

如果前述以太局域网中的工作站发送数据给另一局域网中的工作站，则该工作站中应用所生成的数据单元 P1 首先封装成一个 TCP 包 P2，例如如图 2 所示（前提是传输层所用协议是 TCP，传输控制协议）。之后，TCP 包被封装成 IP 包 P3，该 IP 包进一步封装成以太网帧 P4，通过局域网发送给连接该局域网的路由器，该路由器还具有一个面向 ATM 网络的接口。在 ATM 适配层中，路由器删去以太网段，将 IP 包分解成 ATM 信元。需要注意的是，封装可以包括在包之前或之后插入（所谓的尾部(trailer)）。

图 3 说明了一个 IP 包 30 的结构（即 IP 数据报）。包头最小为 20 字节，它被划分成 5 个 4 字节“字”，在该图中以连续行形式表示。该图在每个域名之后的括号中给出了以比特为单位的域长。消息头首先包括 4 比特版本域 31，表明所用的 IP 版本。之后是长度域 32（IHL，因特网消息头长度），表明消息头以 4 字节字为单位的长度。类型域 33 表明业务类型，域 34 是包括消息头的数据报总长。标识域 35 用以在包重构时标识该 IP 包。标志域 36 使得系统能够推断出分解包的片断（fragment）是否是原包的最新片断。片断偏置域 37 的内容则表明该片断在原包中处于哪个位置。生存期域 38 表明包可以在网络中存在的最长时间。包所经过的每个路由器都需要评估该域的值。协议域 39 表明该数据报所承载数据的高层协议（例如 TCP）。域 40 包含一个消息头校验和。域 41 和 42 用于源和目的地址，即它们以 32 比特地址的形式指示发送主机和接收主机的地址。地址域之后是可选域 43，它很少用到。该域中传送的数据一般涉及网络测试或故障修正；该数据可以例如定义数据包应当经过的给定路由。在需

要的时候，可以在该域中增加填充比特，使得字节的数量可以被 4 整除。

在上述消息头之后才是 IP 包的 actual 数据。数据域的长度可以变化，但其上限由域 34 的长度限定，这意味着整个包的最大长度是 2^{16} 字节。

前面说过，在按照图 1 的 ATM 网络中，IP 包以 ATM 信元的形式传送。图 4a 示出了 ATM 网络中需要传送的一个信元的基本结构。该网络中需要传送的每个信元包括一个 48 字节的净荷和 5 字节消息头，而消息头的精确结构（消息头的内容）与不同情况下使用 ATM 网络的那一部分相关，因为 ATM 网络体系结构包括在标准中详细指定的接口数量，ATM 信元中采用的消息头结构与所使用的接口（即网络的哪部分）相关。

图 4b 示出了 ATM 网络的 UNI（用户-网络接口）接口上的信元消息头结构，它是 ATM 终端设备和 ATM 节点之间的接口。图 4c 则说明了 ATM 网络的 NNI 接口（网络-网络接口）上的信元消息头结构，它是两个 ATM 节点之间的接口，这两个 ATM 节点可以位于该网络中，也可以分处两个网络。

信元消息头的路由寻址域包括虚路径标识符（VPI）和虚信道标识符（VCI）。在图 4b 所示的消息头结构中，总共为路由寻址域（VPI/VCI）保留了 24 比特，因此它仅用于用户终端。在图 4c 所示的消息头结构中，为路由寻址域（VPI/VCI）保留了 28 比特，它用于 ATM 网络的所有其它部分。如其名所示，路由寻址域是 ATM 网络中信元进行路由寻址的基础。虚路径标识符 VPI 主要用于网络内部，它实际上经常决定信元应当寻路到哪个物理连接。而虚信道标识符 VCI 则通常仅用于在网络边界的路由寻址。但是应当注意，只有 VPI 和 VCI 相结合才能明确定义信元的路由。

在规范中定义的 ATM 信元消息头的其它域是

GFC（通用流量控制），用于在用户终端进行业务量监控的域，尚未精确定义；

PTI（净荷类型指示符），主要用于区分网管信元和用户信息信元，此外根据是否在该路由上检测到拥塞，可以进一步区分用户信息信元；

CLI（信元丢弃优先级），用于针对信元丢弃概率定义信元的优先级（主要对应于帧中继网络中 DE 比特）；

HEC（消息头差错控制），消息头校验和

在这些域中，仅有 PTI 涉及本发明。该域可以用于，例如传送与虚路径和虚信道相关的拥塞消息。但是，本发明采用 PTI 域的另一特性，下面将详细予以描述。

通常到达 ATM 适配层的是不同格式的信号，ATM 适配层的任务一方面是，在将这些信号发送到 ATM 网络之前，将其转换成 ATM 网络所需的标准格式，另一方面是，根据从 ATM 网络到达的信元重构这些信号，然后进一步转发给用户或控制接口。为不同业务类别（A...D）标准化了不同类型的适配层（AAL1...AAL5）。例如 AAL 类型 3、4 和 5 为源和目的之间没有时间相关性的应用提供传输业务。

图 5 说明了 AAL 5 进行的将 IP 包分段成 ATM 信元，以及根据 ATM 信元重构包，它说明了图 1 所示的示例性网络中 ATM 适配层的操作。ATM 适配层通常划分成两个子层，它们由缩写 SAR（分段和重组子层）以及 CS（会聚子层）表示。CS 子层完成用户数据单元（例如 IP 包）和控制数据的封装/去封装。CS 子层进行封装之后得到的帧被称为 CS-PDU（会聚子层协议数据单元）。AAL 5 的封装使得迹部分被加入用户数据单元（例如 IP 包），包含例如差错检查部分（CRC）。迹部分的长度为 8 个字节。整个 CS-PDU 的长度对应于 48 比特的倍数，通过适当地增加填充域 PAD 可以实现这一点，PAD 在包的迹部分和净荷之间具有 0..47 字节长度。

SAR 子层在传输方向上将每个 CS-PDU 分段成 48 字节的字段，称为 SAR-PDU（分段和重组子层协议数据单元）。在接收方向上，将 SAR-PDU 组合在一起生成了 CS-PDU。

ATM 适配层之下的 ATM 层负责在需要发送的 SAR-PDU 中加入

5 字节消息头域 CH (图 4b 和 4c), 从而生成 ATM 信元 50, 发送到 ATM 网络。ATM 网络仅处理信元头, ATM 网络中不处理, 甚至不读 48 字节的净荷。在接收方向上, ATM 层将消息头从信元中删去, 将 48 字节净荷提供给 SAR 子层以进行组装。

在图 1 网络中, 路由器发送 IP 包, 它们以前述方式将包分段成信元, 将包发送到 ATM 传输链路。链路反向一侧的路由器以前述方式根据信元重构包, 以传统方式基于 IP 地址进行路由选址, 将包再次分段成信元, 在下一 ATM 链路上传输。

正常情况下, 包的路由寻址由软件决定。如果软件实现的路由寻址和前述包分段和重构相组合, 则基于 ATM 的路由器网络的操作将大大减慢。这种传统路由寻址方法的实现也相当昂贵。

为了缓解这些缺陷, 已开发出一种称为 IP 交换的方法。IP 交换基于流概念: 流 (通常) 是从同一源发往同一目的的一系列 IP 包。因此, 一个流 (通常) 包括源和目的地址相同的 IP 包。例如 TCP 连接是一个流: 当打开 TCP 连接时, 从源向目的发送一系列包。在 IP 交换中, 路由器节点识别流并在网络边缘请求路由器给每个流的包配上一个唯一的流标识, 例如一个唯一的 VPI/VCI 标识符。当属于给定流的包配备了唯一的 VPI/VCI 标识符时, 网络内的路由器能够利用通常的 ATM 交换, 在信元层实现包转发。因此, 路由寻址不需要进入第 3 层 (IP 层), 而可以在第 2 层 (ATM 层) 实现。

但是, 以上现有技术方案的缺陷在于, 它预先假定所有路由器都能够识别流并前转流信息。已知的路由器无法进行这样的流控制。此外, 该方案将在网络中产生额外的业务量, 因为路由器必须请求发送端路由器将流标识符插入包。

本发明目的是消除以上缺陷并提供一种方案, 该方案改进了路由器容量, 而不要求应用任何新的流量控制协议。

该目的通过独立的权利要求书所定义的方案来实现。

本发明的思想在于, 基于下层协议数据单元监控较高协议层上两个连续包之间的边界, 在每个不同包开始时, 从下层协议数据单元中

读取高层地址（包地址）。据此对包的第一数据单元进行路由选择，之后只要下一数据单元中包括的下层连接标识符域第一数据单元相同，则相同的路由寻址自动应用到该数据单元。只要具有相同连接标识符的数据单元内容表明高层包仍在进行，则继续自动路由选择。

在本发明的优选实施例中，下层数据单元是 ATM 信元；因此，本发明首先采用 ATM 适配层的特性，按照 ATM 适配层，信元消息头中的 1 个比特表明是否是根据该包生成的最后信元。该比特是 PTI 域中的第 3 比特。本发明所采用的另一属性是 AAL5 帧没有可以利用单个 VPI/VCI 标识符进行信元复用的流标识域。因此，可以确信，通过同一虚连接传送的，在指示包结束的信元之前的所有信元都属于同一包。

由于按照本发明的方案不需要分段和重构包，所以路由器的吞吐量得以增加。这可以通过传统路由器实现。因此，网络中不需要附加的流控制协议，从而路由器也较为简单，不会引起网络的额外业务量。

下面通过附图的例子，结合图 6 和 7 详细描述本发明及其优选实施例，在附图中

图 1 示出了可以应用本发明的环境；

图 2 说明了在包被发送到网络之前，包的生成；

图 3 示出了 IP 包的结构；

图 4a..4c 示出了 ATM 信元的一般结构；

图 5 说明了 ATM 适配层 5 所完成的将包分段成 ATM 信元，以及根据 ATM 信元重构包；

图 6 的流程图说明了本发明的方法；以及

图 7 是实现本发明方法的路由器的功能结构框图。

在按照本发明的方案中，IP 包的分段在所述网络边缘的节点中进行，ATM 信元被发送到 ATM 传输链路。

如前所述，PTI 域的最后 1 个比特（图 4a..4c 中的比特号 2）表明何时开始一个新的高层包（IP 包）。如果该比特被置成 1，则是 IP

包的最后信元，在这种情况下，下一包将以下一非空闲信元开始。用户数据（非控制数据）的 PTI 域具有值 000 或 010，则该信元是包中的第一信元或者包中的一个信元。另一方面，用户数据的 PTI 域具有值 001 或 011，则该信元是包中的最后信元。

如果 ATM 网络中的路由器接收 ATM 信元，则可以基于 PTI 域的最后一个比特发现 IP 包的第一信元。按照本发明的路由器利用了该属性。下面结合图 6 予以描述，图 6 说明了本发明的路由器操作。

开始时路由器监控通过 VPI/VCI 连接到来的 ATM 信元中 PTI 域的第 3 比特值（步骤 61）。只要该比特值保持为 0，则继续进行监控。一旦检测到任一虚连接的该比特已被置成 1，则立即假定通过该虚连接的下一非空信元为包的第一信元。因此，该方法可以开始读下一信元。在该信元到达之后，读取其 IP 目的地址（步骤 63）；从图 3 中可以看出，该地址被映射成字段，使得它的位置总是固定在信元的净荷字节 13-17。根据读取的目的地址，从路由表中找出与该地址相关联的输出端口的标识符（步骤 64），将该信元发送到该输出口（步骤 65）。之后，通过所述虚连接监控信元的 PTI 域，只要 PTI 域的第 3 比特保持为 0，即只要包还在进行，则将所有信元都发送到同一端口（步骤 66）。如果检测到 PTI 域中的第 3 比特变成了 1，则该方法在步骤 62 中继续读取下一非空信元。

图 7 说明了路由器中包交换的一种可能实现。该路由器在输入 IN 和输出 OUT 之间包括 3 个连续部件 A、B 和 C。部件 A 定义了与包地址相关联的输出端口，部件 B 完成到正确端口的路由寻址。部件 C 由输出端口组成，其中可以例如为下一链路生成一个新的 VPI/VCI 标识符。

为了临时存储到来信元，在路由器的输入配置了一个信元缓冲器 70；该缓冲器读取到来信元的地址（VPI/VCI 标识符和 PTI 域）。如果信元缓冲器向 IP 地址检索部件 72 指示了包的第一信元，则检索部件从信元中读取 IP 地址，将该地址传送到内容可编址存储器（CAM）73。内容可编址存储器也称为关联存储器；它是一个存储位置基于其

内容识别的存储器。存储器 73 根据该 IP 地址，将与该 IP 地址相关联的输出端口的地址返回给检索部件。检索部件将指示输出端口的标识符传送给配置处理 71，后者据此定义一个可用的出 VPI/VCI 标识符，将输出端口和出 VPI/VCI 标识符的数据写入 ATM 交换机 76（在对应该输入端口和入 VPI/VCI 标识符的行中）的路由表（表 T1）。每个包写一次，因此属于同一包（即具有相同 VPI/VCI 和相同输入端口）的所有信元被交换到同一输出端口，并接收同一出 VPI/VCI 对。

只要信元缓冲器 70 发现该包仍在继续，并且包的 VPI/VCI 标识符与该包中的第一信元相同，它就将这些信元直接传送到 ATM 交换机，后者基于其路由表完成路由选择。这样，在每个包开始时配置 ATM 交换机，之后该包中的所有信元自动切换到正确的输出端口。

为了加快处理，内容可编址存储器用作主路由表 75 的高速缓存。如果存储器 73 中没有发现 IP 目的地址，则信元可以缓存，通过传统软件检索处理 74 直接从主路由表 75 搜索输出端口。内容可编址存储器基于检索更新。

利用以上方法，可以大大加快包的路由寻址，路由器的配置也达到最简。

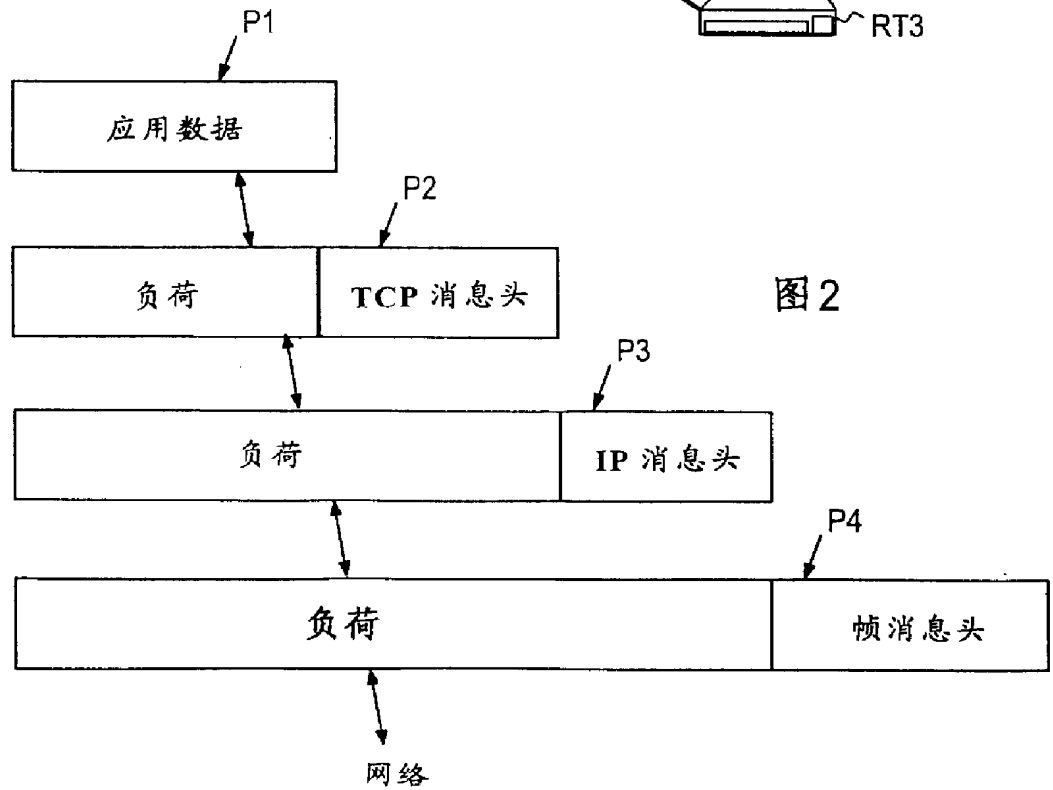
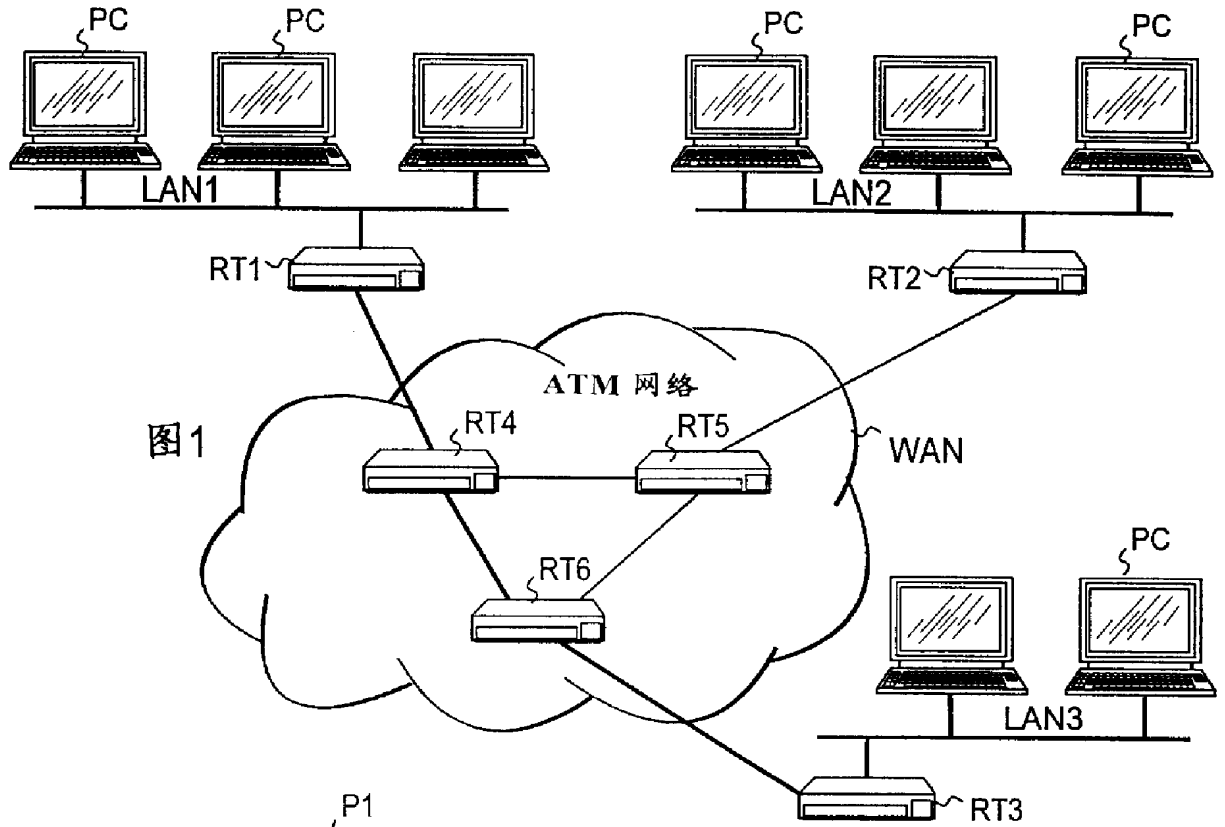
以上提出的例子都与网络层所用协议为 IP 的情况相关，这是该层所用协议的最常见情况。但是，按照本发明的方案并不局限于 IP，也同样适用于包具有标准格式地址，从而可以在下层数据单元中发现该地址的任一其它（网络层或更高层）协议。尽管按照本发明的方案主要用于网络传送 ATM 信元，但该方案并不局限于此，下层可以采用任一协议，只要在该协议数据单元中能够识别较高协议层数据单元（包）间的边界。

原则上也可能在第一数据单元中无法立即发现上层地址（取决于数据包中该地址的位置），而必须存储该第一数据单元，直至发现了上层地址。

尽管以上结合按照附图的例子解释了本发明，但需要理解，本发

明并不局限于此，在此后附权利要求书所提出的创新思想范围内可以进行改进。

说明书附图



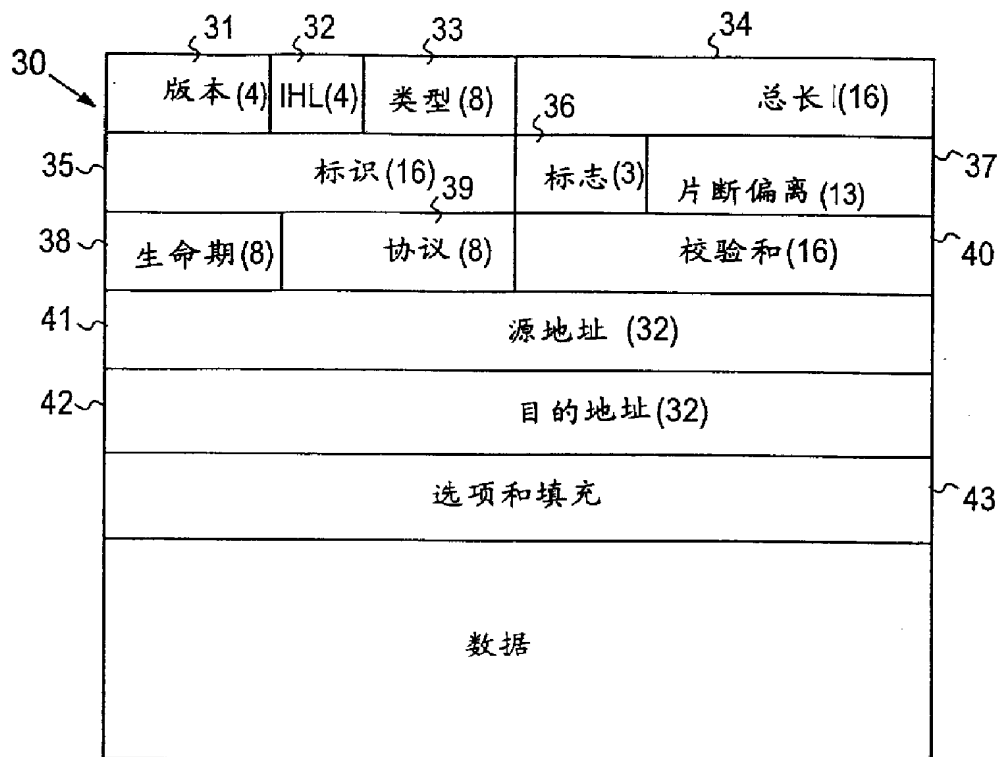
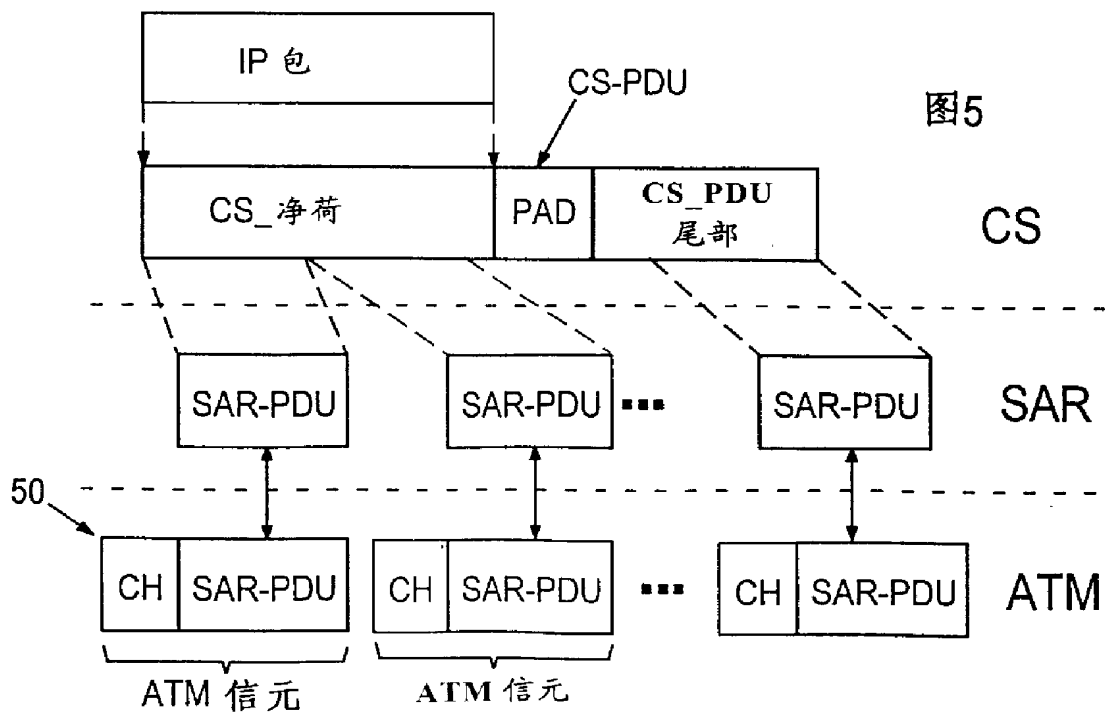


图3



比特： 8 7 6 5 4 3 2 1

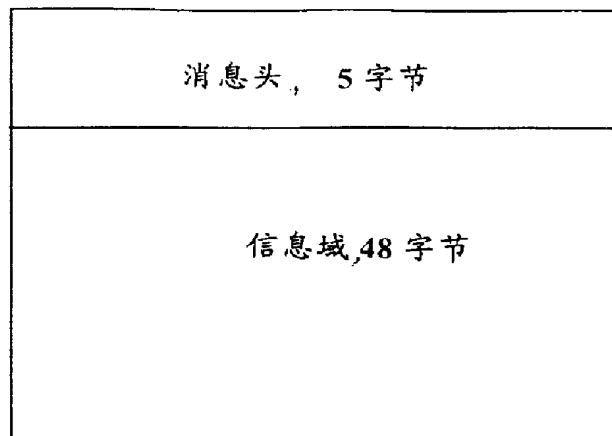


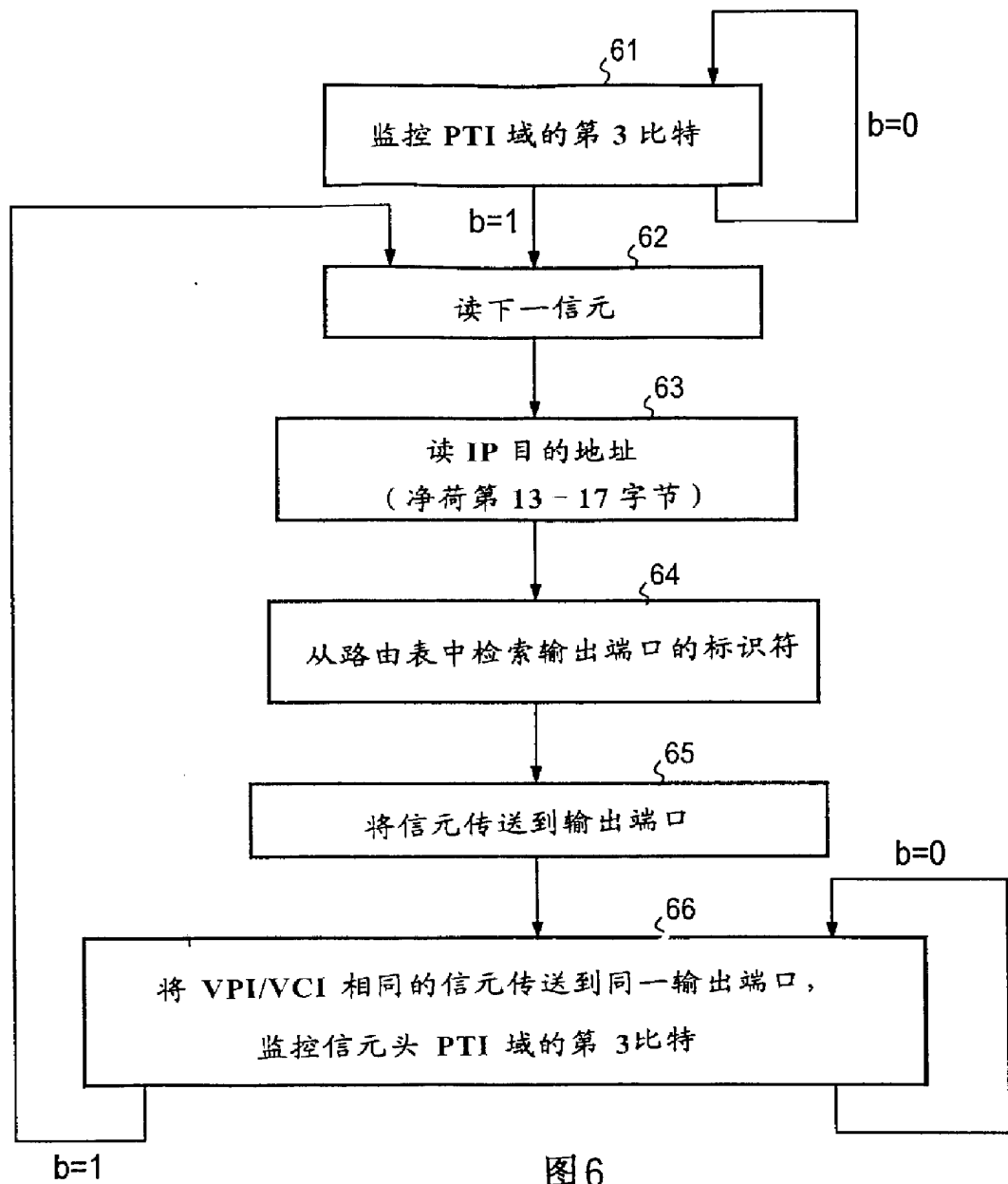
图4a

8	7	6	5	4	3	2	1	BIT / BYTE
GFC				VPI				1
VPI				VCI				2
VCI								3
VCI				PTI		CLP		4
HEC								5

图4b

8	7	6	5	4	3	2	1	BIT / BYTE
VPI								1
VPI				VCI				2
VCI								3
VCI				PTI			CLP	4
HEC								5

图4c



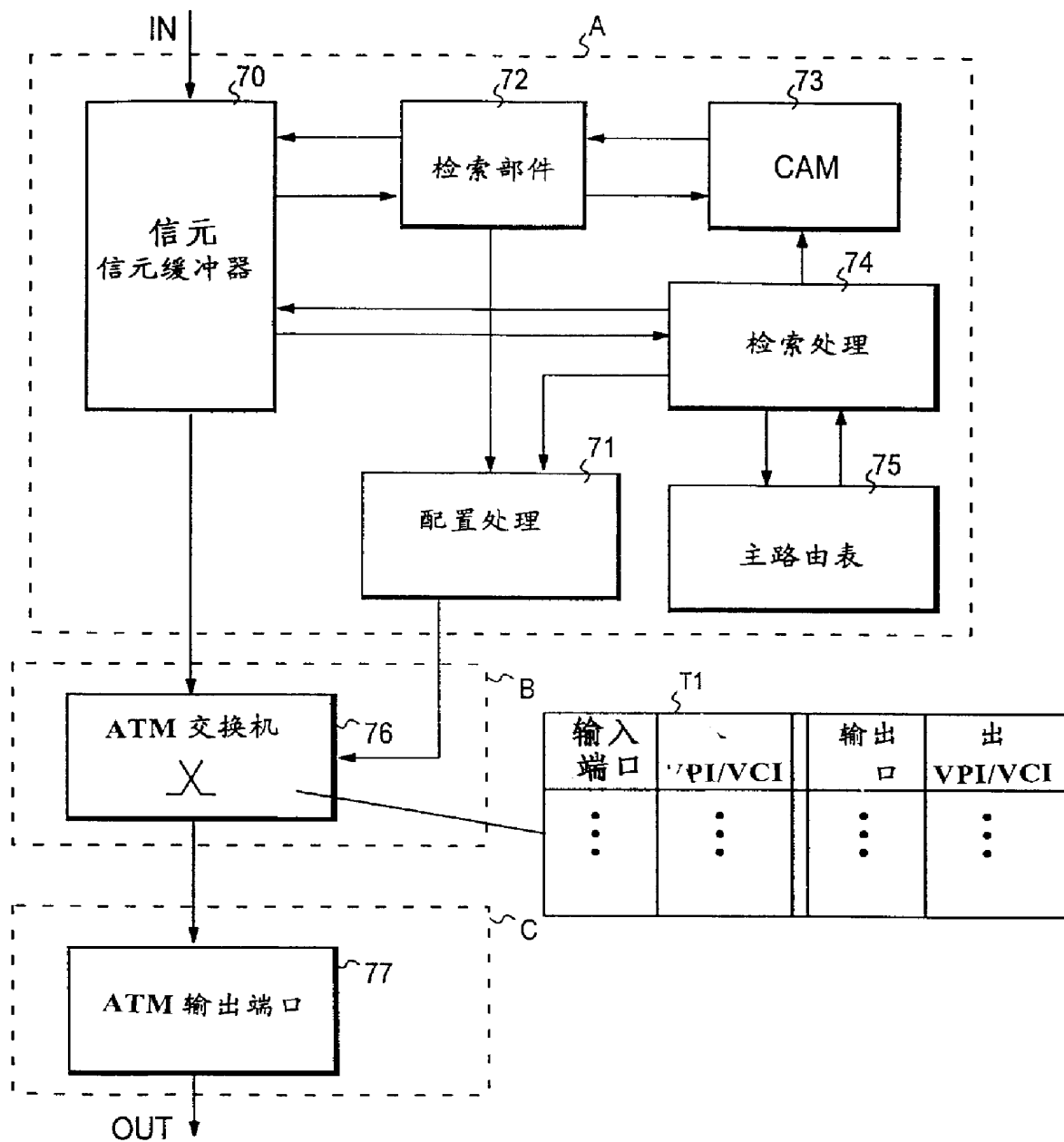


图 7